Auditory neuroscience: **Is speech special?**

David R. Moore

Speech is thought to be perceived and processed in a unique way by the auditory system of the brain. A recent study has provided evidence that a part of the brain's temporal lobe is specifically responsive to speech and other vocal stimuli.

Address: University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, UK.
E-mail: dmoore@physiol.ox.ac.uk

Current Biology 2000, 10:R362–R364

0960-9822/00/$ – see front matter

A regular feature of the late-night, British television chat show *So Graham Norton* is that the host searches the Web for providers of dubious 'personal services'. He then telephones the hapless provider. The highly embarrassing conversation that follows is broadcast to the studio audience and to the home viewer. To the delight of the audience, the interviewee has been known to recognise Norton's voice, and the way in which (s)he responds to this surprisingly welcome intrusion adds to the fun.

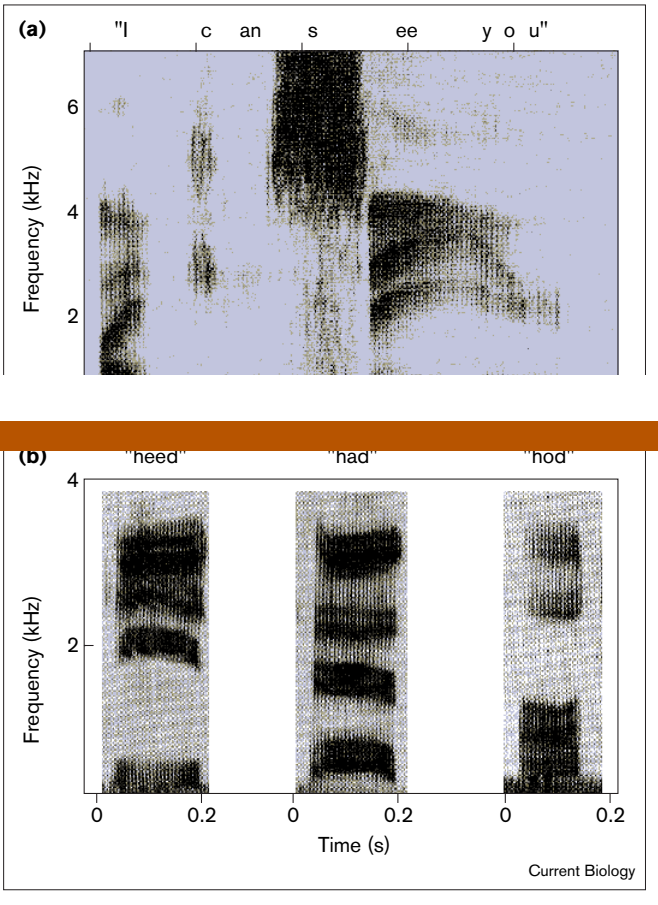
The recognition of an unexpected voice under these circumstances is a stunning ability, made more so by the effortless-

phone network and masked by the background studio sounds. They must decipher the rapid and simultaneous fluctuations in time, frequency and level of several streams of information (Figure 1). Finally, they must choose between hundreds or thousands of potential speakers, each carrying their unique acoustic signature, yet each speaking the same identifiable and interpretable language. The notion that this capability reflects the special way that speech is processed by the auditory system has received support from the recent discovery that part of the brain's temporal lobe is specifically responsive to speech and other vocal stimuli [1].

Speech is special

The complexity of speech perception has led some [2] to suggest that speech is 'special', that it utilises a special mode of hearing and that it engages unique neural mechanisms. There are several ways in which speech perception appears to differ from other types of hearing [3]. An important one is the variable nature of the acoustic cues producing a given percept. Speech sounds also tend to blend together. For example, there may not be a pause between spoken words, but there will be pauses during other parts of the speech stream. In addition, the acoustic character of a consonant is often dictated by the character of the vowel that follows it (Figure 1). Another way in which speech differs from non-speech sounds is through categorical perception. Normally, it is easier to discriminate between two sounds along a continuum than it is to identify the individual sounds. With speech, the opposite is true. Some very small changes in the acoustic properties of a signal can lead to a shift in the identity of a speech token whereas other, larger changes will make no perceptible difference. For these reasons, general speech recognition devices have been very difficult to design.

Figure 1



Speech spectrograms. **(a)** A spectrogram of a male talker saying "I can see you". The intensity of the speech is shown by the darkness of the shading. Most of the speech (except /s/) is grouped into dark frequency bands called formants. There are gaps of low intensity within the speech, but not necessarily between words. (reproduced from [3] with permission.) **(b)** Three instances, spoken by a British male, of the variable spectral representation of the consonant /h/, as used in the words "heed, had, hod". Although we clearly hear the /h/ sound, its spectrum is heavily influenced and largely masked by the following vowel. (Adapted from [12]).

Neural processing of speech

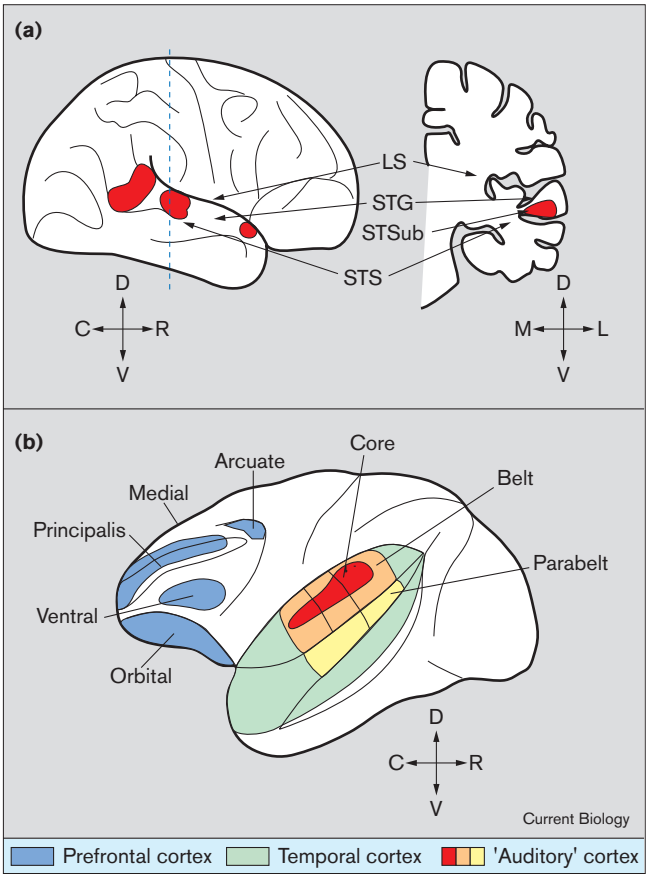
It has long been suspected that the brain handles speech in a different way to other acoustic signals. In the 1950s, experiments using 'dichotic' listening, where different signals are presented to each ear, showed a right ear advantage for the detection and tracking of speech, whereas the reverse was true for melodies [3]. As each ear excites neurons preferentially in the contralateral auditory cortex, these findings supported the suggestion, originally based on studies of brain lesioned patients, that the left cerebral cortex is specialised for speech. A recent paper by Belin and coworkers [1] has now provided evidence that neurons in a region of the temporal cortex, the upper bank of the superior temporal sulcus, are selectively activated by voices, both speech and non-speech (Figure 2).

Using functional magnetic resonance imaging (fMRI), Berlin *et al.* [1] showed in three experiments that: first, the superior temporal sulcus was, bilaterally, more activated by a cocktail of vocal sounds (speech and non-speech) than by an energy-matched cocktail of non-vocal sounds (such as animal cries or 'mechanical sounds'); second, the superior temporal sulcus was more activated by vocal sounds than by a variety of non-vocal sounds of human origin, including scrambled voices that preserved the energy envelope of vocal sounds but that did not perceptually resemble normal, vocal sounds; and third, that activation of the superior temporal sulcus by bandpass-filtered vocalizations and non-vocalizations correlated with the ability of listeners to perform a vocal/non-vocal decision task and a speaker gender identification task using the same stimuli. All stimuli produced a higher activation of the left cortex, but the superiority of the vocal stimuli was more marked in the right cortex. In both hemispheres, vocal superiority was most marked in the caudal superior temporal sulcus.

Cortical pathways for sound processing

These results [1] suggest that the caudal superior temporal sulcus may play a pivotal role in the processing of vocal signals by the brain. It seems anatomically well placed to do so as, in the macaque monkey, its neurons receive extensive input from other, apparently lower-order auditory areas in the parabelt region of the superior temporal gyrus, and then project on to the prefrontal cortex [4,5]. It has been proposed [5,6] that connections passing from the belt and, particularly, the parabelt regions of the superior temporal gyrus to the prefrontal cortex may form part of distinct circuits that separately process sound location and identification (Figure 2). A caudal circuit, targeting the caudal, dorsal and lateral prefrontal cortex, is proposed to form part of a 'where' (location) pathway, whereas a rostral and ventral prefrontal cortex area is the target of a 'what' (identity) pathway. The proposed functional segregation relies heavily on analogies with the visual system, and on physiological findings of neurons responsive to vocal and spatially restricted stimuli in the prefrontal cortex.

Figure 2



Putative cortical pathways for vocal processing. **(a)** Voice selective areas in the upper bank of the human superior temporal sulcus (STSub), as revealed by fMRI. Areas shaded in red are those shown to be significantly more activated (BOLD) by vocal than by non-vocal stimuli. The left drawing shows the lateral (sagittal) aspect of the right cerebral cortex. The right drawing shows a frontal (coronal) slice through the right cortex at the level indicated by the blue line in the left drawing. Note that, in the left drawing, the activation actually lies within the upper bank of the superior temporal sulcus rather than on the surface of the superior temporal gyrus. Abbreviations: C, caudal; D, dorsal; L, lateral; M, medial; R, rostral; V, ventral; LS, lateral sulcus; STG, superior temporal gyrus; STS, superior temporal sulcus. (Results from [1].) **(b)** Auditory areas in the macaque monkey left cerebral cortex. In the temporal lobe, the primary auditory cortex (core, red) is surrounded by and projects to the belt areas (orange). These, in turn, project to the parabelt areas (yellow) and, directly, to the prefrontal areas (arcuate, medial, principalis, ventral and orbital; blue). The parabelt areas project to the superior temporal sulcus (green) and to the prefrontal cortex. Note that all projections within the temporal lobe map caudal–caudal and rostral–rostral. Caudal temporal lobe areas project to the arcuate and principalis, whereas rostral temporal lobe projects to the orbital, ventral and medial prefrontal cortex. (Adapted from [4].)

Belin *et al.* [1] have argued that their results are consistent with participation of the voice-sensitive area of the superior temporal sulcus in the 'what' pathway, as might be expected from the nature of vocal selectivity. In fact, as detailed above, the caudal region of maximal vocal

selectivity within the superior temporal sulcus found in their study appears to fit anatomically into the proposed 'where' stream. This discordance may be due to differences between human and macaque cortical connectivity, and further anatomical evidence is required. In the meantime, it remains an open question as to how the human superior temporal sulcus might relate to functionally homologous areas in the prefrontal cortex and elsewhere.

More generally, the evidence for functionally distinct 'where' and 'what' auditory cortical pathways is, in my view, highly questionable. For example, in support of the proposed 'where' system, physiological studies of the macaque prefrontal cortex [7,8] have been cited. These studies have shown single neuron spatial receptive fields that all tend to be centred on the contralateral pole of auditory space, in keeping with spatial receptive fields in many other areas of the auditory system. Such an organization is not, however, generally considered to represent sound source location. For that purpose, a 'map' of auditory space, as found in the mammalian superior colliculus [9] and barn owl midbrain [10] is the standard. Moreover, Vaadia *et al.* [8] found that only a minority (36%) of dorsal prefrontal cortex neurons had *any* auditory spatial selectivity, and 12% of ventral prefrontal cortex neurons had similar selectivity. This seems thin evidence on which to claim a specific spatial function for the dorsal prefrontal cortex.

Use of fMRI in auditory research

As anyone who has visited an MRI facility will know, the scanner produces intense (>100 dB SPL), complex sounds during rapid image acquisition. Because auditory cortex neurons respond non-linearly to sounds, it is not advisable to follow the traditional fMRI procedure of subtracting activation during a 'no stimulus' period from that during a 'stimulus' period, where the scanner noise is present throughout. Some researchers [11] have got around this problem by using 'sparse' temporal sampling. This method takes advantage of the slow (longer than four second) rise time of the blood oxygenation level dependent (BOLD) signal acquired in fMRI by interleaving 'quiet' stimulus periods between 'noisy' acquisition periods. The method is relatively inefficient, but does produce images that are much less influenced by scanner noise contamination than the traditional method. The use of a sparse imaging method probably contributed to the success of Belin *et al.* [1] in achieving sensitive differentiation between vocal and non-vocal signals.

Vocal and speech selectivity

A further intriguing aspect of the work reported by Belin *et al.* [1] is that they identified in the brain a voice-sensitive area rather than a speech-sensitive area. The range of stimuli classed as 'vocal' included laughs, sighs and coughs, in addition to isolated words and connected speech in several languages. As data were not presented separately for speech and non-speech stimuli, we do not know whether the superior temporal sulcus is more-or-less selective for these highly varied vocal elements. It is a pity that this distinction was not made, as non-speech stimuli obviously constitute a qualitatively different class of stimuli to speech. In particular, it is the speech part of vocalizations that distinguishes humans from other animals, and it is interesting that animal vocalizations were among the non-vocal stimuli in this study. One can only wonder whether and how the spectro-temporal structure of non-speech human vocalizations provides greater saliency for neurons in the superior temporal sulcus than that of animal vocalizations, which are known to contain many of the characteristics of human speech. Similarly, the mixing of non-native language stimuli with those of the listener's native tongue might obscure potentially interesting distinctions, both in the level and site of activation.

Despite these short comings, the study of Belin *et al.* [1] is an important one, pointing the way with generally appropriate methods to a new and fascinating area of research.

Acknowledgements

My own research is supported by the Medical Research Council, the National Health Service, Defeating Deafness, the John Ellerman Foundation, and MindWeavers Ltd. I would like to thank Andrew King and Jan Schnupp for valuable comments on the paper.

References

1. Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B: **Voice-selective areas in human auditory cortex.** *Nature* 2000, **403:**309-312.
2. Liberman AM: **On finding that speech is special.** *Am Psychologist* 1982, **37:**148-167.
3. Moore BCJ: *An Introduction to the Psychology of Hearing*. 4th edn. San Diego: Academic Press; 1997.
4. Kaas JH, Hackett TA, Tramo MJ: **Auditory processing in primate cerebral cortex.** *Curr Opin Neurobiol* 1999, **9:**164-170.
5. Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP: **Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex.** *Nat Neurosci* 1999, **2:**1131-1136.
6. Kaas JH, Hackett TA: **'What' and 'where' processing in auditory cortex.** *Nat Neurosci* 1999, **2:**1045-1047.
7. Azuma M, Suzuki H: **Properties and distributions of auditory neurons in the dorsolateral prefrontal cortex of the alert monkey.** *Brain Res* 1984, **298:**343-346.
8. Vaadia E, Benson DA, Heinz RD, Goldstein MH: **Unit study of monkey frontal cortex: active localization of auditory and visual stimuli.** *J Neurophysiol* 1986, **56:**934-952.
9. King AJ, Hutchings ME: **Spatial response properties of acoustically responsive neurons in the superior colliculus of the ferret: a map of auditory space.** *J Neurophysiol* 1987, **57:**596-624.
10. Knudsen EI, Konishi M: **A neural map of auditory space in the owl.** *Science* 1978, **200:**795-797.
11. Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW: **'Sparse' temporal sampling in auditory fMRI.** *Hum Brain Mapp* 1999, **7:**213-223.
12. Ladefoged P: *A Course in Phonetics.* 3rd edn. Fort Worth: Harcourt Brace; 1993.